



**Universidade de Brasília - UnB
Faculdade de Ciência da Informação - FCI
Curso de Graduação em Biblioteconomia**

**WEB SEMÂNTICA E FOLKSONOMIA NAS REDES SOCIAIS:
RELAÇÕES E APLICAÇÕES NA CIÊNCIA DA INFORMAÇÃO**

Daiane Sorato

Brasília, julho de 2011

Daiane Sorato

**WEB SEMÂNTICA E FOLKSONOMIA NAS REDES SOCIAIS:
RELAÇÕES E APLICAÇÕES NA CIÊNCIA DA INFORMAÇÃO**

Monografia apresentada a Faculdade de Ciência da Informação, como requisito para obtenção do grau de Bacharel em Biblioteconomia.

Orientadora: Prof^a. Dr^a. Marisa Bräscher Basilio Medeiros

Brasília, julho de 2011



Título: Web semântica e folksonomia nas redes sociais: relações e aplicações na ciência da informação.

Aluna: Daiane Conceição Sorato.

Monografia apresentada à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Brasília, 13 de julho de 2011.

Marisa Brascher Basilio Medeiros – Orientadora

Professora da Faculdade de Ciência da Informação (UnB)

Doutora em Ciência da Informação

Ilza Leite de Azevedo Santos Lopes - Membro

Professora da Faculdade de Ciência da Informação (UnB)

Doutora em Ciência da Informação

Eliana Carlan – Membro externo

Consultora Técnica do Ministério da Saúde

Área de Disseminação da Informação

Aos meus pais, Cida e Ladislau,
por serem a minha razão de viver.
Com muito amor e carinho.

Agradecimentos

À professora Marisa, minha orientadora, pela atenção, incentivo e pela importante contribuição acadêmica.

A todos os professores da FCI, pelos ensinamentos.

Aos colegas de trabalho que, apesar das adversidades, sempre me incentivaram e torceram pelo meu sucesso.

Aos meus amigos, por estarem ao meu lado nos inúmeros momentos de incertezas, desabafos, frustrações e alegrias, e por todo o carinho e atenção que sempre me dispensaram.

Ao Eduardo, meu namorado e companheiro de aventuras, por me apoiar em todas as minhas (in)decisões e me fazer tão feliz.

Aos meus irmãos, Derik e Geovana, por existirem e encherem a minha vida de alegria. Aos meus pais, por me mostrarem a importância dos estudos e da dedicação ao trabalho, sempre com muito amor, paciência, dedicação e serenidade.

A todos os que estão ao meu lado nesse momento tão importante da minha vida.

Muito obrigada!

“Tornamos nosso mundo significativo pela coragem de nossas perguntas e pela profundidade de nossas respostas”.

Carl Sagan

Resumo

Discute as conseqüências do crescimento informacional e a necessidade de tratamento adequado da informação. Com base na literatura disponível, aponta as técnicas e ferramentas em desenvolvimento que visam otimizar o uso e recuperação da informação na web, como os metadados, as ontologias e a Web Semântica. Destacam-se as aplicações das tecnologias da Web Semântica e das folksonomias das redes sociais. Estas ferramentas são vistas no contexto dos ambientes informacionais, como as bibliotecas digitais. Destaca-se a importância do envolvimento do profissional da Ciência da Informação com as tecnologias de informação e comunicação.

Palavras-chave:

web semântica; metadados; ontologias; bibliotecas digitais; redes sociais; folksonomia.

Abstract

Discusses the consequences of informational growth and the necessity of a proper treatment of information. Based on the literature available, points out the techniques and tools in development that aim to optimize use and recovery of information in the web, like metadata, ontologies and semantic web. It is highlighted the applications of semantic web technologies and folksonomies of social networks in information environments, like digital libraries, and the importance of the involvement of the information science professional with technologies of information and communication.

Keywords:

semantic web; metadata; ontologies; digital libraries; social networks; folksonomy.

Lista de figuras

Figura 1 – Representação da Cauda Longa na Folksonomia.....	28
Figura 2 – Representação Iterativa (visão detalhada).....	38

Lista de abreviaturas e siglas

CGM	Consumer-Generated Media
CML	Chemical Markup Language
HTML	Hyper Text Markup Language
IMRAD	Introduction, Method, Results and Discussion
MathML	Mathematical Markup Language
OCA	Ontologia do Conteúdo de Conhecimento em Artigos Científicos
OWL	Web Ontology Language
RDF	Resource Description Framework
RSS	Really Simple Syndication
SBML	System Biology Markup Language
SRI	Sistema de Recuperação da Informação
STMML	Scientific Technical and Medical Markup Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
XML	eXtensible Markup Language
W3C	World Wide Web Consortium
Web	World Wide Web
www	World Wide Web

Sumário

1 Introdução.....	12
2 Problema.....	14
3 Objetivo Geral.....	15
3.1 Objetivos Específicos.....	15
4 Justificativa.....	16
5 Metodologia.....	17
6 Metadados.....	18
6.1 A iniciativa Dublin Core.....	19
7 Ontologias.....	20
8 Web Semântica.....	22
8.1 XML e RDF.....	23
9 Redes sociais.....	24
9.1 A Folksonomia nas redes sociais.....	26
9.2 Web Semântica e tratamento da informação nas redes sociais.....	29
10 Aplicações da Web Semântica e o uso de folksonomias no contexto da Ciência da Informação.....	31

10.1 Ontologias em bibliotecas digitais.....	33
10.2 A revolução dos motores de busca.....	34
10.3 Folksonomia na Ciência da Informação.....	35
10.3.1 Representação Iterativa e Folksonomia Assistida.....	37
11 Considerações finais.....	40
12 Referências.....	42

1 Introdução

A quantidade crescente e desordenada de informação que se encontra disponível em meio digital na Internet é uma realidade a qual todos devem se adaptar. Vários problemas surgem em decorrência dessa realidade, entre eles a deficiência na representação dos recursos digitais que possam conduzir a uma recuperação de documentos que sejam de fato relevantes ao usuário. Além da quantidade, a variedade de formatos, como música, vídeos, imagens e animações, extrapolam as técnicas clássicas de recuperação de informações baseadas principalmente em formatos textuais (GALDO; VIEIRA; RODRIGUES, 2009).

Segundo Souza e Alvarenga (2004), apesar de ter sido projetada para facilitar o acesso à informação, a web foi implementada de forma descentralizada e quase anárquica, crescendo de maneira exponencial e caótica, deixando a desejar quando surge a necessidade de recuperação de informação em meio ao imenso repositório de documentos na qual a Internet se tornou.

A indexação dos conteúdos na web, em geral, não é complementada pela semântica, motivo pelo qual o trabalho de selecionar os documentos que correspondem ao contexto correto da busca e descartar informações inúteis ainda é do ser humano. Os mecanismos de busca existentes na atual web não conseguem fazer o tratamento semântico do que é requisitado, gerando contratempos relacionados à polissemia e sinonímia, dentre outros. A solução para esse problema está na implementação da chamada Web Semântica.

O uso das folksonomias, por meio da atribuição de etiquetas (*tags*) por parte de usuários das redes sociais, apresenta-se como um trampolim para a aplicação do tratamento semântico da informação em recursos digitais, na medida em que podem ser criadas e/ou aproveitadas as técnicas utilizadas e promovidas pelos desenvolvedores dessas redes.

Nesse contexto, este trabalho baseia-se em uma pesquisa bibliográfica nas áreas de Web Semântica, redes sociais, folksonomia (*tagging* e *social tagging*), motores de busca e bibliotecas digitais, com o objetivo de explicitar a influência dessas áreas no desenvolvimento atual e no futuro da Ciência da Informação.

Este documento está dividido em 4 (quatro) partes. Na primeira, apresentam-se os conceitos de metadados e ontologias, para em seguida ser explicada a estrutura da Web Semântica. Em seguida, abordam-se os conceitos de redes sociais, bem como o papel da folksonomia no tratamento da informação nas redes sociais. Na terceira parte, discute-se a ligação entre a web semântica e o tratamento da informação nas redes sociais. Por fim, aponta-se a importância dessa ligação para a Ciência da Informação, como no desenvolvimento de motores de busca mais eficientes e o tratamento da informação nas bibliotecas digitais.

2 Problema

O problema do tratamento da informação na web é uma realidade e traz importantes questionamentos para os profissionais que lidam com a informação. Com o advento da Web 2.0, a interatividade tornou-se um fenômeno predominante na Internet, e a intensa participação do usuário na criação dos conteúdos disponíveis na rede pode ter aspectos negativos ou positivos, no que diz respeito à representação dos recursos¹.

A Web Semântica apresenta uma solução alternativa para indexação de recursos digitais, que atribui significado às etiquetas e propõe uma melhor interação entre humanos e máquinas. Porém, essa web ainda está em desenvolvimento e a literatura sobre o assunto, na área de Ciência da Informação, em língua portuguesa, é escassa.

As possibilidades de integração entre o uso da folksonomia nas redes sociais no contexto do desenvolvimento da Web Semântica, abrindo portas para um novo paradigma no tratamento da informação, foram as principais motivações para o desenvolvimento deste trabalho.

¹ Entende-se recurso como qualquer objeto que pode ser identificado com uma URL.

3 Objetivo Geral

Produzir um texto, baseado em revisão de literatura, que apresente as relações existentes entre a Web Semântica, representação da informação nas redes sociais e a importância desse processo no contexto da Ciência da Informação.

3.1 Objetivos Específicos

1. Efetuar levantamento bibliográfico sobre Web Semântica, redes sociais, folksonomias, motores de busca, bibliotecas digitais.
2. Identificar e sintetizar as abordagens dos diferentes autores sobre os assuntos abordados.
3. Produzir um texto que apresente as relações existentes entre a Web Semântica, as redes sociais e possíveis aplicações para a Ciência da Informação.
4. Contribuir para disponibilizar literatura sobre o tema em língua portuguesa para consulta por alunos de graduação.

4 Justificativa

Os estudos sobre a Web Semântica são relativamente novos, pelo fato da web 3.0 ainda se apresentar em fase de desenvolvimento. A sua importância para a Ciência da Informação já foi descrita em outros trabalhos, e será reforçada na presente monografia.

Observa-se, ainda, que o modo como a informação é tratada pelos usuários da web 2.0 (nas redes sociais), por meio da folksonomia e do uso de *tags*, pode ser uma importante contribuição no desenvolvimento das tecnologias necessárias para que os objetivos da Web Semântica sejam atingidos.

Além disso, a partir dos estudos sobre o comportamento do usuário na identificação de recursos digitais nas redes sociais, será possível a criação de motores de busca cada vez mais eficientes, que tragam resultados de fato relevantes no momento da pesquisa.

Dessa forma, pretende-se com o presente trabalho contribuir para um maior entendimento sobre a Web Semântica e o crescente uso das folksonomias nas redes sociais, enfatizando a aplicação dessas tecnologias na área de Ciência da Informação, como desenvolvimento de motores de busca semânticos e tratamento da informação em bibliotecas digitais, e disponibilizando literatura em língua portuguesa sobre o tema.

5 Metodologia

A metodologia utilizada para a realização deste trabalho foi a revisão de literatura nas áreas de Web Semântica, folksonomia, redes sociais, os novos motores de busca e os projetos de bibliotecas digitais com tratamento semântico. Na coleta dos dados foram utilizadas as palavras-chave “web semântica”, “redes sociais”, “folksonomia”, “tags”, “motores de busca”, “bibliotecas digitais”, além de várias combinações entre os termos. A pesquisa bibliográfica foi realizada no período de dezembro de 2010 a junho de 2011, nas seguintes fontes:

Sites:

Google Acadêmico: www.scholar.google.com

W3C World Wide Web Consortium - www.w3.org

Artigos:

Ciência da Informação – Revista do IBICT

DataGramaZero – Revista de Ciência da Informação

Em Questão – Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS

Encontros Bibli – Revista Eletrônica de Biblioteconomia e Ciência da Informação

Informação & Informação – Revista da Universidade Estadual de Londrina

Informação & Sociedade: Estudos – Revista da Universidade Federal da Paraíba

Journal of Digital Information

Liinc em Revista – Laboratório Interdisciplinar em Informação e Conhecimento

PontodeAcesso – Revista do Instituto de Ciência da Informação da UFBA

Perspectivas em Ciência da Informação – Revista da UFMG

TransInformação – Revista da PUC de Campinas

6 Metadados

Na Ciência da Informação, os metadados vêm sendo utilizados no tratamento da informação, sendo formas de representação de recursos informacionais para fins de descrição, identificação, localização, busca e recuperação. Para Rocha (2004, p. 113),

o conceito de metadados é bastante simples: metadados são dados sobre dados. Quando se trata do mundo digital, chama-se de recurso o objeto descrito por metadados, pois este pode ser tanto um simples dado, quanto um documento, uma página da web, ou até mesmo uma pessoa, uma coleção, um sistema, um equipamento ou uma organização. Na web, o conceito recurso significa qualquer objeto que pode ser atingido através de uma URI (*Uniform Resource Identifier* - www.w3.org/Addressing), como qualquer recurso que é acessado via seu endereço eletrônico. Isso inclui documentos, páginas pessoais, sites ou sistemas. A descrição de uma pessoa ou organização é feita através da descrição da página dessa pessoa ou organização.

O objetivo do uso de metadados no contexto da web não é somente descrever os recursos eletrônicos para a avaliação de relevância por usuários humanos, mas também possibilitar que agentes inteligentes (programas especiais, robôs e agentes de *software*) compreendam os metadados associados aos documentos e possam recuperá-los, avaliar a sua importância e manipulá-los com mais eficiência (CASTRO; SANTOS, 2007).

Rocha (2004) diz que os metadados podem estar inseridos nos próprios documentos que descrevem. Em páginas HTML (*Hyper Text Markup Language*), pode-se inserir metadados através de marcações (*tags*) que não são visíveis no navegador, mas podem ser lidas por motores de busca e outros aplicativos que os usam. Normalmente, esquemas de metadados são desenvolvidos por comunidades especialistas em um domínio do conhecimento. Nos casos em que existe a necessidade de utilização de um esquema que descreva recursos que abrangem vários domínios do conhecimento, seriam combinados e adaptados vários elementos de diversos esquemas de metadados.

6.1 A Iniciativa Dublin Core

Para facilitar esse procedimento, a Iniciativa Dublin Core² oferece uma coleção básica de elementos de metadados, para serem utilizados por qualquer comunidade, independente da sua área de domínio. O esquema Dublin Core foi o resultado de uma oficina realizada em 1995, patrocinada pela comunidade da Ciência da Informação e por especialistas em padrões para a web, cuja tarefa principal concentrou-se em identificar e definir um conjunto de elementos de metadados, denominado Dublin Core, para descrever recursos da web, a fim de proporcionar a descoberta destes recursos (WEIBEL *apud* ROCHA, 2004).

Os elementos do Dublin Core foram criados com o propósito de serem simples, de fácil manipulação e aplicáveis a vários domínios. Estão organizados em três grupos: conteúdo, propriedade intelectual e instância. Como esse esquema foi criado tanto para quem gostaria de construir uma *home page* e descrevê-la de forma genérica, quanto por especialistas de um domínio do conhecimento, foi criado o conceito de “qualificadores” (Dublin Core *Qualifiers*), onde os recursos de um domínio específico são descritos pelos mesmos elementos do Dublin Core (ROCHA, 2004).

Ainda segundo o mesmo autor, esquemas de metadados como o Dublin Core são muito úteis para a Web Semântica, pois podem ser uma boa base para a inserção de elementos semânticos na descrição de recursos informacionais. Porém, técnicas de raciocínio e inferência precisam ser incorporados para que as máquinas possam auxiliar os humanos no uso da informação descrita, o que já vem sendo realizado pela Web Semântica, na medida em que estende o conceito de metadados.

² <http://dublincore.org>

7. Ontologias

O termo ontologia, adotado no contexto da Ciência da Computação com sentido diverso ao da filosofia (ciência ou estudo do “ente” ou “ser”), é definido como sendo a descrição de um determinado domínio da realidade, através de termos e relações (descrição formal dos conceitos), com o objetivo de criar um vocabulário compartilhado para troca de informações entre membros de uma comunidade, sejam humanos ou agentes inteligentes. O uso das ontologias é essencial para a construção de mecanismos que visam à interoperabilidade entre sistemas heterogêneos (construídos por diferentes fabricantes e adotando tecnologias diversas) e à implementação das tecnologias da Web Semântica. Segundo Sure e Studer (*apud* SAYÃO; MARCONDES, 2008),

as ontologias estabelecem uma compreensão compartilhada de um domínio de interesse para apoiar a comunicação entre seres humanos e agentes computacionais. As ontologias são representadas caracteristicamente por uma linguagem de representação processada por computador, sendo considerada uma tecnologia-chave para o desenvolvimento da Web Semântica.

De acordo com o SemanticWeb.org, “uma ontologia é uma especificação de uma conceituação. É designada com o propósito de habilitar o compartilhamento e reuso de conhecimentos, de forma a criar ‘compromissos ontológicos’, ou definições necessárias à criação de um vocabulário comum”.

Segundo Bräscher (2002), a ontologia na Web Semântica possui uma taxonomia e um conjunto de regras de inferência. A taxonomia define classes de objetos e relações entre eles. Pode-se expressar um grande número de relações entre entidades assinalando propriedades para as classes e permitindo que subclasses herdem suas propriedades. O conjunto de regras de inferência confere um poder adicional às ontologias.

As ontologias diferem-se dos tesauros. Os tesauros, apesar de servirem como instrumentos de construção de vocabulários controlados para determinar quais os termos que podem ser usados e estabelecem entre eles relações de equivalência (sinonímia), hierarquia e associação, são menos flexíveis e complexos que as

ontologias, que permitem que novos relacionamentos sejam estabelecidos. Segundo Pickler (2007), “as ontologias podem ser redefinidas ou adaptadas em determinados contextos, e suas relações são extensíveis, conforme necessário”. Para Boccato, Ramalho e Fujita (*apud* Dziekaniak, 2010),

há relacionamentos que são possíveis nas ontologias e não os são em tesouros, como por exemplo, as relações inversas ‘se uma pessoa é_dona de um cão, logo o cão tem_dono’ e as relações transitivas ‘se Joaquina é uma praia de Santa Catarina e Santa Catarina fica no Brasil’, então a ontologia infere, e passa a admitir que, ‘a praia de Joaquina fica no Brasil’ – mesmo sem precisarmos estabelecer essa constatação.

Para Brandão e Lucena (*apud* Dziekaniak, 2010), além de definir conceitos provenientes de diversos domínios do conhecimento, evitando ambigüidades, o uso de ontologias possibilita a definição de relacionamentos entre esses conceitos e, principalmente, a realização de inferências, propriedade que faz das ontologias a melhor solução para sistemas de recuperação de informação.

Na Web Semântica, o uso das ontologias permite que os documentos sejam representados por termos relacionados entre si, o que permite que a informação neles contida seja processada por programas que “entendem” o sentido do conteúdo descrito. Para que as ontologias possam ser aplicadas no contexto da Web Semântica, o W3C³ (*World Wide Web Consortium*) propõe o padrão OWL (*Web Ontology Language*), que possui maior facilidade para expressar a semântica das terminologias usadas para descrever os documentos da web. O padrão OWL baseia-se na linguagem XML (*eXtensible Markup Language*), que será explicada mais adiante.

³ <http://www.w3c.org> – “O *World Wide Web Consortium* (W3C) é um consórcio internacional no qual organizações filiadas, uma equipe em tempo integral e o público trabalham juntos para desenvolver padrões para a Web. Liderado pelo inventor da web Tim Berners-Lee e o CEO Jeffrey Jaffe, o W3C tem como missão conduzir a World Wide Web para que atinja todo seu potencial, desenvolvendo protocolos e diretrizes que garantam seu crescimento de longo prazo”.

8. Web Semântica

Os estudos sobre a Web Semântica são relativamente novos e estão em fase de crescimento. A Web Semântica vem sendo desenvolvida e será uma “extensão da web atual. Nela a informação é dada com um significado bem definido, permitindo melhor interação entre os computadores e as pessoas” (BERNERS-LEE *apud* SOUZA; ALVARENGA, 2004). Ainda de acordo com o autor, nesse novo contexto as máquinas terão acesso a coleções estruturadas de informações (dados e metadados) e de conjunto de regras de inferência que ajudarão no processo de dedução e na administração do raciocínio automatizado, ou seja, a representação do conhecimento. Propõe-se, dessa maneira, a criação de mecanismos que possibilitem uma maior interoperabilidade entre sistemas de informação heterogêneos, a fim de que passem a adotar ferramentas comuns de representação da informação, como esquemas de metadados, classificações, tesauros e ontologias. Com isso, a tarefa de interpretar o sentido de uma busca realizada pelo usuário passará a ser, também, das máquinas que serão capazes de compreender a informação representada semanticamente.

Para atingir esse objetivo, é necessária a padronização de tecnologias, linguagens e metadados descritivos, a fim de que os usuários sigam regras comuns e compartilhadas sobre como armazenar dados e descrever a informação, podendo esta ser utilizada por outros usuários, humanos ou não, de maneira automática e não ambígua (SOUZA; ALVARENGA, 2004).

O objetivo da Web Semântica é estabelecer novos padrões de descrição de recursos digitais através da atribuição de semântica, o que possibilitará a compreensão, por parte dos agentes inteligentes, do significado dos documentos recuperados, e assim promover a manipulação inteligente da informação e a disponibilização de dados úteis ao usuário. A tarefa de selecionar os resultados relevantes de uma pesquisa, antes atribuída ao ser humano, é assim delegada aos computadores, que terão a capacidade de processar essas informações. Segundo Hendler (*apud* DZIEKANIAK; KIRINUS, 2004),

a Web Semântica pode ser considerada como a composição de um grande número de pequenos componentes ontológicos que apontam

entre si. Dessa forma, companhias, universidades, agências governamentais e grupos de interesses específicos procurarão ter seus recursos web ligados a um conteúdo ontológico, já que ferramentas poderosas serão disponibilizadas para intercambiar e processar essas informações entre aplicações web.

Pereira e Cruz (2010) definem Web Semântica como sendo um conjunto de instrumentos que fazem parte da web 2.0. Para eles, o objetivo da Web Semântica é permitir que máquinas entendam o significado dos recursos descritos culturalmente. “*Tags, posts, blogs, wikis* são instrumentos desta ‘web social’, que possibilitam a comunicação e a organização coletiva, fazendo a interface com mecanismos complexos, controlados e formais de representação computacional da informação” (PEREIRA; CRUZ, 2010). A web 2.0, bem como os instrumentos descritos acima, serão objetos de estudo mais adiante.

8.1. XML e RDF

Para que os computadores possam entender o conteúdo da web e ler os dados estruturados e conjuntos de regras que os ajudam a conduzir seus raciocínios, as páginas devem ser escritas em uma nova linguagem.

Segundo Dziekaniak e Kirinus (2004), dentre as tecnologias desenvolvidas para a Web Semântica está o XML (*eXtensible Markup Language*), linguagem de marcação que permite aos usuários criarem tags personalizadas sobre o documento criado, diferentemente do HTML, que possui estrutura de tags fixas, impedindo a criação de novos tipos de descritores. Segundo Jardim e Palazzo (2009), a linguagem XML é uma das tecnologias mais amplamente adotadas para intercâmbio e representação de informações na web, pois permite, a cada um, criar suas próprias *tags*, rótulos ocultos que anotam as páginas web, ou seções de texto em uma página.

Existem experiências baseadas em linguagem XML para marcação e publicação de artigos científicos na web. Marcondes (2011), citando vários autores, apresenta propostas como CML (*Chemical Markup Language*), SBML (*System*

Biology Markup Language), MathML (*Mathematical Markup Language*), além de outras com enfoques mais gerais como STMML (*Scientific Technical and Medical Markup Language*).

A tecnologia RDF (*Resource Description Framework*) é uma aplicação XML que serve como uma base para o processamento de metadados (JARDIM; PALAZZO, 2009). O RDF trabalha com um trio de informação o qual expressa o significado das informações. Cada componente do trio tem sua própria finalidade, em analogia ao sujeito, verbo e objeto de uma frase e recebe uma identificação URI (*Uniform Resource Identifier*). Exemplo de URI é a URL (*Uniform Resource Locator*).

Tanto o XML como o RDF são recomendados pelo W3C (*World Wide Web Consortium*) para gerar linguagens de marcação na construção de sites da web.

9. Redes sociais

Anterior à chamada web 3.0, como é conhecida a Web Semântica, a web 2.0 está presente no cotidiano dos usuários da rede mundial de computadores. É caracterizada principalmente pela intensa participação do usuário, seja pela possibilidade de criação e compartilhamento de conteúdos por qualquer pessoa, profissional ou não, seja pela interação do usuário com serviços on-line oferecidos por empresas, por exemplo. Segundo O'Reilly (2006 *apud* JARDIM; PALAZZO, 2009), a regra mais importante na web 2.0 “é desenvolver aplicativos que aproveitem os efeitos de rede para se tornarem melhores quanto mais são usados pelas pessoas, aproveitando a inteligência coletiva”.

O termo web 2.0 foi proposto entre 2003 e 2004, e foi pensada inicialmente como uma estratégia de *marketing* segmentado. No entanto, passou a dar ênfase à efetiva colaboração do usuário no processo de organização da informação “e proporcionou, desse modo, alteração das lógicas de organização da informação, agora não mais centralizadas no papel do mediador da informação” (MOURA, 2009, p. 3). Para a autora:

O funcionamento da Internet e os diversos produtos e serviços a ela vinculados alteraram significativamente os padrões de agregação social na contemporaneidade. A relativização das noções de tempo e espaço e a redução dos rituais sincrônicos abriram espaço para a mobilidade e o estabelecimento de comunidades não constrangidas pela dimensão geográfica e ocasionou também a implementação de novos padrões de cooperação. A difusão global de informações permitiu uma série de agregações que se constituem em torno do interesse informacional, tornado fluxo. É nesse contexto que surgem as comunidades virtuais, uma modalidade de agregação de sujeitos dispersos geograficamente em torno de interesses comuns.

É nesse contexto que surgem as redes sociais. Conforme Dias e Silveira (2007 *apud* MOURA, 2009), as redes sociais “caracterizam-se por processos produtivos e de experiência, poder e cultura, dinamismo e descentralização na tomada de decisão, autonomia dos membros, horizontalidade das relações e desconcentração do poder”. Segundo Cardozo (2008 *apud* JARDIM; PALAZZO, 2009), “a questão central das redes é a valorização dos elos informais e das relações, em detrimento das estruturas hierárquicas”.

Ao compreender o potencial de interação das redes sociais, os provedores de soluções colaborativas em rede incorporaram funcionalidades aos *softwares* para facilitar a cooperação a partir do pacto de linguagem, movimento esse que tem causado “enorme impacto sobre a organização da informação em ambientes digitais” ao criar uma grande expectativa de que as redes e seus membros seriam capazes, eles próprios, “de coordenar os fluxos informacionais e o provisionamento de soluções no âmbito da gestão dos sistemas de informação” (MOURA, 2009).

São disponibilizadas inúmeras ferramentas, aplicativos e *softwares* que proporcionam interoperabilidade de serviços, como *wikis*, *blogs*, *fotologs*, *videologs* e fóruns de discussão, onde está presente a CGM (*Consumer-Generated Media* ou mídia gerada pelo consumidor), além de aplicativos para compartilhamento de *links*, RSS (*Really Simple Syndication*) e as folksonomias.

9.1 A Folksonomia nas redes sociais

A necessidade de organização da informação na web e o aumento da participação do usuário no processo de criação e organização de recursos formam o cenário para o surgimento da folksonomia. Termo formado a partir da junção das palavras “*folk*” (povo) e “*taxonomy*” (taxonomia) e cunhado pela primeira vez por Thomas Vander Wal em 2004, a folksonomia consiste no uso da linguagem natural no processo de representação da informação por usuários integrantes de contextos sociais participativos, através de etiquetas (*tags*) e/ou palavras-chaves (metadados) associadas aos conteúdos a serem evidenciados, visando à sua recuperação, podendo também agregar valor para a formação de nuvens de indicadores (*tag clouds*) (PEREIRA; CRUZ, 2010). A folksonomia também é conhecida como “etiquetagem social” (*social tagging*), “classificação social” (*ethnoclassification*), “etiquetagem colaborativa” (*collaborative tagging*), *social bookmarking*, taxonomia popular, entre outros.

Segundo Trant (2008), podemos pensar na etiquetagem (*tagging*) como um processo (com foco na escolha da terminologia pelo usuário); folksonomia como o vocabulário coletivo resultante do processo (com foco na organização do conhecimento); e na etiquetagem social (*social tagging*) como um contexto técnico-social no qual a etiquetagem ocorre (com foco na computação e redes sociais).

Pereira e Cruz (2010) consideram que “a classificação das informações através de *tags* é um processo de comunicação que deixa vestígios do comportamento informacional, que por sua vez consolidam práticas culturais e discursivas”. A colaboração espontânea dos usuários por meio das redes sociais e de relacionamento é um exemplo de como as ferramentas tecnológicas se mesclam com fenômenos humanos (GALDO; VIEIRA; RODRIGUES, 2009).

O desenvolvimento das folksonomias reflete a dificuldade e os altos custos da aplicação de vocabulários controlados no contexto da web (SPITERI, 2007). Além disso, a informação, principalmente na web, torna-se obsoleta mais rapidamente, e há uma dificuldade cada vez maior no tratamento do volume de informação disponível. Estudos apontam para a possibilidade de melhorias na indexação

profissional de recursos da web com materiais provenientes de outras fontes, como a exploração do uso e processamento da linguagem natural atribuída por usuários aos recursos digitais.

Com relação ao status de instrumento de representação atribuído às folksonomias, Moura (2009) diz que

(...) em um ambiente em que é crescente a presença dos usuários que conduzem suas pesquisas e de autores que disponibilizam conteúdos on-line sem a intervenção de mediadores, tornou-se evidente que o campo da organização da informação precisaria rever seus métodos com vistas a compreender esse movimento de transformação intelectual inevitável.

De acordo com Pereira e Cruz (2010), a definição de um vocabulário depende de valores culturais, devido ao caráter social das regras estabelecidas entre uma comunidade específica. Assim, os significados das marcações de conteúdos são variados, pois a web coloca em contato comunidades com identidades distintas. As anotações diversas dos conteúdos, permitidas pelas linguagens de marcação, promovem uma interconexão semântica.

As folksonomias são expressões das necessidades informacionais dos usuários das redes sociais onde elas são utilizadas. Essas linhas de expressão também podem refletir necessidades de comunidades de interesse: usuários que usam um mesmo tipo de *tag* formam grupos onde cada um pode fazer pesquisas usando técnicas simples, obtendo resultados satisfatórios (SPITERI, 2007). A folksonomia é fruto da necessidade humana de comunicação, integração e convivência.

Vander Wal (*apud* TRANT, 2008) propõe um estudo dos sistemas de etiquetagem social definindo folksonomias gerais e específicas. Folksonomias gerais (como ocorrem no Delicious) são o resultado de muitas pessoas “etiquetando” o mesmo item: a análise dessas *tags* revela uma lei de distribuição onde itens que interessam a poucas pessoas formam a chamada “cauda longa”.

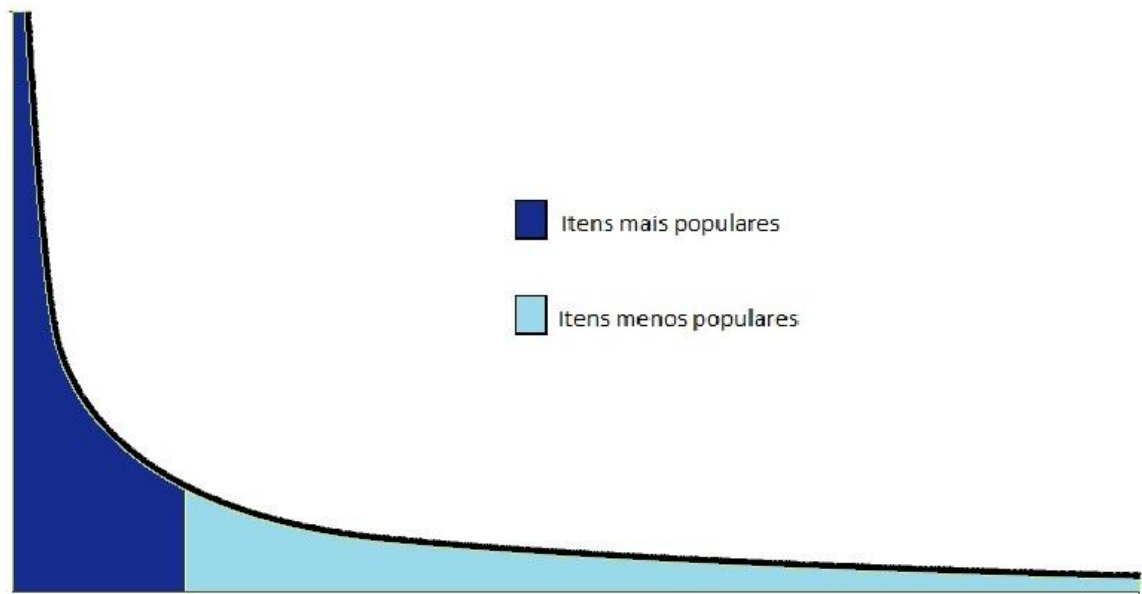


Figura 1 – Representação da cauda longa na folksonomia

A cauda longa é um conceito que se baseia na regra dos 80/20 de Pareto, que sugere que 80% das receitas de uma empresa provém de 20% de seus lucros. No contexto das folksonomias gerais, pode ser interpretada da seguinte maneira: os itens situados na parte azul escura da figura correspondem àqueles que são mais “etiquetados” pelos usuários, seriam os itens mais populares. Já os que se situam na parte azul clara são os itens menos populares.

As folksonomias específicas (como o Flickr⁴) ocorrem quando apenas uma pessoa (ou poucas) etiquetam um objeto, geralmente criado pelo próprio usuário. Nesse contexto mais individual, pode ser mais difícil identificar o relacionamento entre *tags*, pois a linguagem é subjetiva.

⁴ <http://www.flickr.com> – De acordo com informações do próprio site, trata-se de um aplicativo *online* de gerenciamento e compartilhamento de fotos e vídeos, que permite que o usuário disponibilize e organize da maneira que achar mais conveniente os seus recursos.

9.2 Web Semântica e tratamento da informação nas redes sociais

Dix et al. (2006 *apud* TRANT, 2008), em um estudo teórico sobre ontologias emergentes, verificaram que sistemas de etiquetagem colaborativa são um mecanismo alternativo de abordagem para a Web Semântica, onde especialistas podem construir ontologias. Um “halo semântico” de significados é construído em torno da *tag* através de consulta automatizada. Relacionamentos entre *tags*, incluindo termos gerais, termos específicos e termos relacionados, são computados, tendo em conta mudanças temporais na definição. Eles propõem um algoritmo que leve em conta a “agregação” (todas as *tags* unidas por co-ocorrência), “abstração” (*tags* gerais e específicas relacionadas), “ambiente” (contexto) e “idade” (ambiente ordenado ao longo do tempo) (TRANT, 2008).

Segundo Mote (2006 *apud* CARATINO; BAPTISTA, 2007), o termo folksonomia representa *social ontologies*, ou seja, ontologias construídas de forma colaborativa, e significa uma classificação consensual gerada pelos usuários dos recursos digitais.

Iniciativas para promover comunidades de cooperação, visando o aprimoramento tanto do uso compartilhado da informação quanto o *marketing* dirigido às características específicas dos usuários, podem ser observadas em sites como o Delicious⁵ e o Faviki⁶ (MOURA, 2009).

Segundo informações contidas no site da Wikipédia, o Delicious oferece um serviço online que permite ao usuário arquivar e catalogar seus sites preferidos para acessá-los de qualquer lugar. Para isso, são atribuídas *tags* às páginas, permitindo a organização de uma lista personalizada de *bookmark*, que fica disponível *online* e pode ser compartilhada com a rede social. De acordo com Moura (2009),

Ao realizar uma pesquisa com a *tag folksonomy*, por exemplo, chegou-se a uma lista das principais páginas que foram socialmente indexadas utilizando o termo. As páginas são acompanhadas do número de usuários que a indicaram e as demais *tags* atribuídas a essa página. Ao navegar nas páginas hierarquizadas é possível identificar no universo semântico, novas *tags* capazes de representar

⁵ <http://www.delicious.com>

⁶ <http://www.faviki.com>

o conteúdo analisado, bem como identificar sujeitos e grupos interessados no tema.

O Faviki é uma ferramenta cuja proposta é permitir que o usuário retire a ambigüidade de sua lista de *bookmark*, construindo uma ponte entre as *tags* e seus significados, e foi fundado em 2010 por Vuk Milicic, desenvolvedor e web designer que acredita que *tags* semânticas podem ser a solução para a “bagunça” de informações na web. O Faviki trabalha em conjunto com duas iniciativas de colaboração na web, o DBpedia e a Wikipédia, para possibilitar a marcação semântica dos *bookmarks*⁷. Beckett (2006 *apud* TRANT, 2008) já havia proposto um método de uso de espaço *wiki* para a desambiguação e estruturação das *tags* através de seus significados.

Segundo informações do próprio site⁸, a Wikipédia é uma enciclopédia *online* que utiliza a ferramenta *wiki*, a qual permite a edição coletiva de documentos através de um sistema que não exige a revisão do conteúdo antes de sua publicação. Apesar da aparente falta de controle do conteúdo, este se faz necessário e existe em vários níveis. Assim, as regras de utilização do site são estabelecidas pelos seus usuários, com o monitoramento e a participação de todos, possibilitando a manutenção da confiabilidade da enciclopédia.

O DBpedia permite que os usuários visualizem a relação entre uma *tag* e seu significado, extraíndo entradas estruturadas da Wikipédia. Através do Zemanta, um mecanismo de sugestão de conteúdos utilizados por blogs e outros geradores de conteúdo, o Faviki recomenda as *tags* semânticas mais utilizadas, para que haja uma maior padronização dos termos.

Moura (2009) enfatiza, ainda, o uso do aplicativo Google Trends, que tem por objetivo avaliar, por meio de algoritmos, o interesse mundial em determinado assunto. Segundo a autora, de posse desse monitoramento terminológico de grande escala proporcionado pelo aplicativo, o elaborador da linguagem “poderá obter os balizadores para a tomada de decisão sobre a inclusão de um termo em um tesauro ou ontologia e o nível de relação proposto na sintaxe para o mesmo, por exemplo”.

⁷ <http://davidkuhta.com/2010/03/semantic-web-in-action-faviki>

⁸ <http://pt.wikipedia.org/wiki/wikipedia>

10. Aplicações da Web Semântica e o uso de folksonomias no contexto da Ciência da Informação

A necessidade de armazenamento e disponibilização cultural da informação pelo ser humano é antiga. As bibliotecas tradicionais desempenhavam esse papel, tornando o conhecimento acessível localmente, a todos que pudessem dirigir-se até elas. Com os avanços tecnológicos, o computador passou a fazer parte do cotidiano das bibliotecas, auxiliando nos processos de catalogação, indexação e consulta ao acervo. A popularização do acesso a Internet e o crescimento dos documentos em formato digital contribuíram para o surgimento das bibliotecas digitais, o que levou a novas reflexões com relação à transferência dos serviços prestados por uma biblioteca tradicional para o meio eletrônico, além do desenvolvimento de métodos de tratamento dos recursos digitais que visem uma recuperação eficiente.

Diante dos avanços tecnológicos que a Web Semântica traz no ambiente da Internet, pode-se pensar em como a recuperação da informação em ambientes informacionais, como as bibliotecas digitais, pode ser beneficiada com a adaptação dessas tecnologias para o ambiente da Ciência da Informação. No âmbito da interoperabilidade, que “tem sido um dos itens mais críticos para quem pensa no desenvolvimento e operação de sistemas de repositórios e de bibliotecas digitais distribuídos funcionando em rede” (SAYÃO; MARCONDES, 2008), tem-se que a chamada interoperabilidade semântica, segundo os mesmos autores, está relacionada com a semântica das informações obtidas de fontes diversas, e que é solucionada através da adoção de ferramentas comuns de representação da informação, como esquemas de metadados, classificações, tesouros e ontologias. Ainda de acordo com Sayão e Marcondes (2008),

espera-se que tecnologias dotadas de habilidades semânticas possam ajudar os usuários de bibliotecas digitais de várias formas. As pesquisas nessa área, de forma geral, investigam como as tecnologias da Web Semântica podem potencializar as funcionalidades das bibliotecas digitais, especialmente na descoberta de recursos com mais eficiência e efetividade, e no compartilhamento de conhecimento no escopo da comunidade do usuário circunscrita pela biblioteca digital.

Castro e Santos (2009) dizem que, no intuito de agregar semântica à descrição bibliográfica para um tratamento mais efetivo de conteúdos digitais, novas tecnologias baseadas em pesquisas na temática Web Semântica estão sendo desenvolvidas.

Marcondes (2011) propõe um modelo semântico de publicações eletrônicas utilizando artigos científicos biomédicos, que tem por objetivo extrair e representar o conteúdo desses artigos em formato “inteligível” por programas, “de modo a permitir que programas realizem ‘inferências’ sobre esse conhecimento, permitindo processar o conhecimento assim recuperado e processado de forma semanticamente mais rica que os atuais SRIs” (MARCONDES, 2011, p. 7). O autor justifica a escolha da área da Medicina pelo fato de os artigos científicos da área seguirem um “rígido padrão formal em seus textos, com seções definidas segundo o chamado padrão *Introduction, Method, Results and Discussion* (IMRAD)” (MARCONDES, 2011, p. 7), o que facilita a descrição semântica dos conteúdos. O modelo ainda conta com a “interface web de autosubmissão de artigos”, que difere das outras interfaces por solicitar ao autor, além dos metadados bibliográficos usuais utilizados para descrever os artigos, a inclusão de suas conclusões. Marcondes (2011, p. 13) afirma que

(...) através dessa interface, os próprios autores entram num diálogo interativo, respondendo a questões colocadas pela interface. A interface realiza processamento de linguagem natural, tanto em trechos do texto do artigo quanto das respostas do autor. O objetivo da interface web é formatar as afirmações feitas pelo autor – as conclusões do artigo - através do processamento lingüístico do texto informado por este, formatando-as segundo relações semânticas (COSTA 2006, 2008), compostas por um Antecedente, um Tipo-de-relação e um Consequente. As relações assim formatadas são registradas em formato “inteligível” por programas, segundo o modelo de conteúdo OCA [Ontologia do Conteúdo de Conhecimento em Artigos Científicos], sendo instâncias desta ontologia e fazendo parte de um pacote, juntamente com os metadados bibliográficos convencionais e o texto digital do artigo. Esses pacotes podem ser tratados por SRIs semânticos, possibilitando buscas semanticamente mais ricas.

O uso de aplicativos da web 2.0 e de folksonomias também aparecem como opções para o tratamento de recursos digitais.

10.1 Ontologias em bibliotecas digitais

Com relação ao uso das ontologias em ambientes de bibliotecas digitais, Ramalho e Fujita (2008, p. 5) dizem que

a utilização de ontologias em bibliotecas digitais possibilita a identificação de zonas de convergência entre os instrumentos convencionais de representação de informações e os novos métodos emergentes de representação de recursos informacionais no meio digital, constituindo-se como um campo fértil para pesquisas, capaz de engendrar uma nova “constelação” epistemológica no desenvolvimento de bibliotecas digitais, apontando perspectivas inovadoras no âmbito da gestão de recursos informacionais e novas possibilidades no fazer profissional da área de Ciência da Informação.

Exemplo de aplicação das tecnologias desenvolvidas a partir da Web Semântica, o projeto MarcOnt surgiu em um cenário onde era necessário solucionar problemas de descrição e representação dos recursos informacionais em ambiente como as bibliotecas digitais. O MarcOnt Initiative se desenvolveu a partir da Biblioteca Digital Semântica JeromeDL⁹. Após alguns problemas iniciais e estudos realizados pelos pesquisadores da biblioteca, decidiu-se criar uma ontologia bibliográfica que contemplasse os aspectos semânticos de forma e conteúdo e que pudesse atuar de forma mais efetiva em uma biblioteca digital semântica.

Segundo Castro (*apud* CASTRO; SANTOS, 2009), o MarcOnt Initiative

seria um padrão de descrição bibliográfica semântica que contempla forma e conteúdo do recurso informacional, a fim de sua identificação e individualização para leitura e processamento compreensíveis por máquina/computador, utilizando-se de instrumentos e ferramentas tecnológicas subjacentes à Web semântica, podendo ser aplicado no contexto de ambientes informacionais digitais.

Considerando que o MARC21, mundialmente aceito e adotado na Biblioteconomia, não possui um padrão de descrição semântica que contemple a profundidade da descrição, ou ainda, que cubra a extensão requerida ou suportada

⁹ <http://www.jeromedl.org/> - De acordo com informações do site, trata-se de uma biblioteca digital social semântica, que permite a publicação de vários formatos de documentos. Os usuários ainda podem disponibilizar bookmarks e anotações e compartilhar conhecimento por meio de uma rede social.

por ele, o padrão de descrição bibliográfica MarcOnt pode possibilitar uma melhora na explicitação das informações, permitindo sua compreensão por máquinas nos atendimentos das solicitações de buscas realizadas pelo usuário final (CASTRO; SANTOS, 2009).

10.2 A revolução dos motores de busca

O desenvolvimento das tecnologias que dão vida à Web Semântica, como a elaboração de páginas em linguagens que permitem a descrição de significados e a construção das ontologias, tem favorecido a criação de novos motores de busca, capazes de agregar o máximo de informações úteis para responder a questionamentos complexos dos usuários. Essa importante aplicação da Web Semântica tem um enorme potencial para a criação de repositórios e bibliotecas digitais que permitirão a realização de pesquisas mais relevantes, em um espaço de tempo cada vez menor.

Ao incorporar essas novas tecnologias, os motores de busca passam a oferecer não apenas a resposta a questão solicitada, como também apresentar soluções alternativas e complementares. Por isso, algumas dessas ferramentas sugerem uma denominação diferente para seus serviços. É o caso do Tumbup¹⁰, que se apresenta como um motor de recomendação social. Ou seja, além de encontrar a informação solicitada, a ferramenta recomenda filmes, livros, artistas, músicas, lugares e produtos, apoiando-se nas atividades dos usuários da rede social Facebook¹¹. No Facebook, os usuários têm a opção de “curtir” *links*, fotos, comentários de outros usuários e assim por diante, além de permitir a criação de aplicativos e serviços paralelos. Assim, o Tumbup faz um rastreamento de toda a esfera social dos usuários do Facebook baseando-se nas tecnologias da Web Semântica, e faz sugestões que condizem com o gosto do internauta.

¹⁰ <http://tumbup.com>

¹¹ <http://facebook.com>

A proposta do True Knowledge¹² é ser a primeira plataforma baseada em inteligência artificial para responder a questões do mundo inteiro. O objetivo do site é oferecer um novo tipo de experiência de busca de informações, onde os usuários podem fazer as perguntas que desejarem em linguagem natural, como se estivessem conversando com outro ser humano. Ao deparar-se com uma questão nunca vista pelo sistema, ele sugere que o usuário insira esses novos dados na base, o que aponta o caráter colaborativo do mecanismo de busca.

O WolframAlpha Computacional Knowledge Engine¹³ tem como objetivo reunir todo conhecimento sistemático em uma única ferramenta, imediatamente calculado e acessível a todos. O motor introduz uma maneira inovadora de obter conhecimento, não pela pesquisa na web, mas trabalhando com dados objetivos, fazendo cálculos dinâmicos baseados em modelos, métodos e algoritmos conhecidos. O formulário de pesquisa aceita toda e qualquer forma de entrada, e gera resultados úteis com o máximo de clareza. É uma ferramenta muito útil para obtenção de resultados na área de ciências exatas.

Esses são alguns exemplos de como a Web Semântica pode ser aplicada em ferramentas que são de grande utilidade para os profissionais da Ciência da Informação. Nunes e Fileto (2007) apresentam uma arquitetura de aplicação para extração de informação de documentos jurídicos baseada em semântica, por meio do uso de ontologias e anotações semânticas. Segundo os autores, a anotação semântica de um documento descreve o seu conteúdo pela associação a conceitos e instâncias descritos na ontologia.

10.3 Folksonomia na Ciência da Informação

Considerando o cenário atual da web, onde os usuários participam desde a criação e o tratamento de recursos até a recuperação da informação, a folksonomia aparece como uma alternativa para a indexação de conteúdos, por configurar-se

¹² <http://www.trueknowledge.com>

¹³ <http://www.wolframalpha.com>

como um rico complemento aos vocabulários controlados já existentes em repositórios científicos.

Dentre as vantagens do uso das folksonomias como ferramentas para o tratamento da informação está o seu caráter inclusivo, ou seja, elas refletem o vocabulário dos usuários, independentemente de pontos de vista, experiências, vieses, e assim por diante (SPITERI, 2007). Além disso, são dinâmicas: não possuem a limitação dos vocabulários controlados no que diz respeito à velocidade de absorção de novos conceitos, que são facilmente acomodados pelas folksonomias.

As desvantagens do uso das folksonomias também são discutidas na literatura. Diversos autores concordam que o seu uso traz problemas inerentes a todos os vocabulários não-controlados, como ambigüidade, polissemia e sinonímia. Sozinhas, elas não fornecem diretrizes para a escolha e a formação de *tags*, como o uso de palavras compostas, pontuação e ordem das palavras (SPITERI, 2007).

Portanto, para que as folksonomias tenham utilidade na atividade de indexação de conteúdos, são necessárias medidas e adaptações que permitam o aproveitamento de todos os aspectos que as tornem vantajosas. Um exemplo de aplicação de etiquetagem social é o projeto Steve Museum¹⁴, que está investigando a possibilidade de uso de folksonomias como novos meios de descrição e acesso à coleções de heranças culturais, por meio do desenvolvimento de ferramentas de *softwares* abertos para a etiquetagem de itens de museus e discussões com membros da comunidade interessados em implementar a etiquetagem social em suas próprias coleções.

Segundo Spiteri (2006), outra forma de uso potencial das folksonomias são os catálogos de bibliotecas públicas, onde os usuários poderiam organizar e etiquetar itens de interesse em pastas específicas, podendo decidir quais *tags* serão públicas.

¹⁴ <http://www.steve.museum>

10.3.1 Representação Iterativa e Folksonomia Assistida

Em estudo para a construção de um modelo para a aplicação da Web Semântica no ambiente de repositórios digitais, Santarém Segundo e Vidotti (2011) apresentam um conjunto de técnicas chamado Representação Iterativa. De acordo com Santarém Segundo (2010), essas técnicas visam

(...) transformar um repositório digital científico em uma ferramenta apta a descrever, armazenar e recuperar informação, permitindo a recuperação semântica e a construção coletiva de uma estrutura relacional semântica de informações por meio de Folksonomia Assistida; e as técnicas utilizadas no desenvolvimento da estrutura sugerida.

O modelo de Representação Iterativa sugere que, na interface padrão que o usuário utiliza para a entrada de dados no repositório digital, haverá uma alteração no campo das palavras-chave, transformando os dados ali inseridos também em *tags* e permitindo a inclusão do conceito de folksonomia no sistema. Isso torna possível a recuperação da informação por meio de nuvens de *tags*, por exemplo (SANTARÉM SEGUNDO; VIDOTTI, 2011).

A Folksonomia Assistida, segundo os mesmos autores, consiste na sugestão de termos que melhor representem o recurso incluído no sistema, apresentados ao usuário no momento da digitação das *tags*, a exemplo do que já é realizado no site Delicious. Esse conjunto de informações sugeridas baseia-se nas *tags* já inseridas no sistema e também nos termos de uma estrutura de representação do conhecimento das áreas de especialidades, que deverá estar associada ao repositório como parte do modelo estrutural proposto. Em um segundo passo, o repositório recebe os dados dos usuários para proceder à pesquisa de relacionamento na estrutura de representação do conhecimento, que no modelo em questão é caracterizada por uma ontologia, que será construída em linguagem OWL (*Web Ontology Language*). Após a pesquisa na estrutura e dentro do conjunto de *tags* já definidas por outros usuários, o sistema faz associações, relaciona termos, e devolve ao usuário um novo conjunto de informações, para a escolha definitiva das *tags* que serão utilizadas. A confirmação dos dados alimenta o sistema, contribuindo para o enriquecimento e fortalecimento do conjunto de termos e relações existentes,

e as informações que foram inseridas tornam-se disponíveis para novos usuários ao efetuarem novos depósitos.

O processo de Representação Iterativa, por meio da Folksonomia Assistida, está representado na figura a seguir.

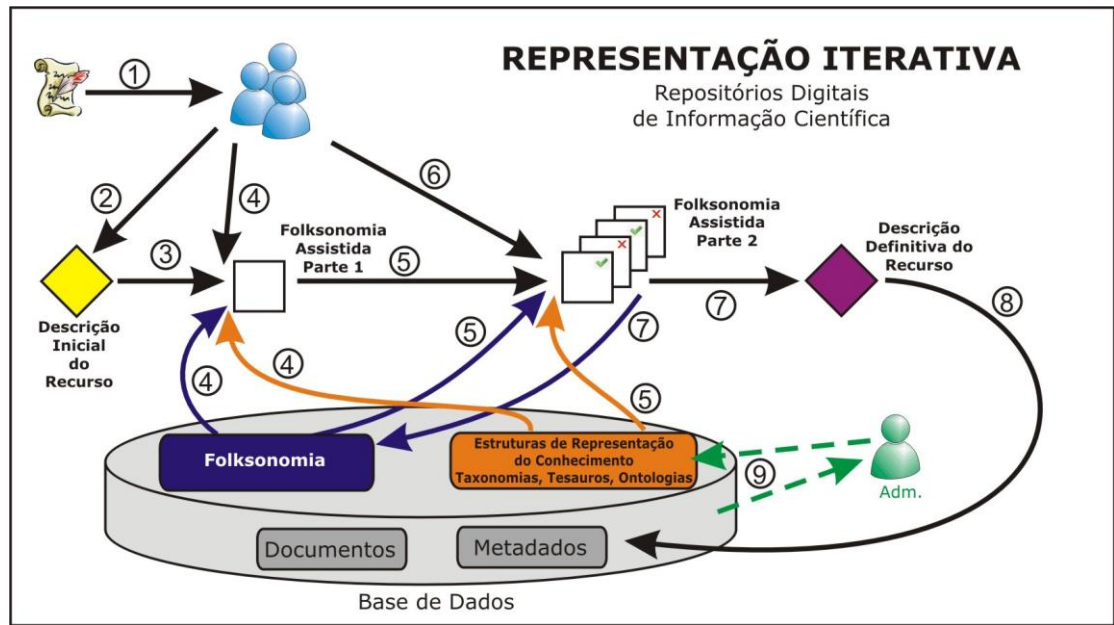


Figura 2 – Representação Iterativa (visão detalhada)
Fonte: Santarém Segundo (2010)

Em primeiro lugar, os usuários fazem a leitura e a verificação do documento a ser depositado, extraindo os metadados necessários, que melhor representem o objeto. Em seguida, inicia-se o processo de descrição inicial do recurso, com exceção da *tag* assunto. É no terceiro passo, com a alimentação da base com as informações dos metadados, que o sistema encaminha o usuário para fazer a descrição da *tag* assunto, momento de maior importância dentro da Representação Iterativa. O quarto passo representa o início da Folksonomia Assistida, onde o sistema colabora na descrição da *tag*, utilizando uma estrutura de sugestão, baseando-se nos termos já inseridos anteriormente por outros usuários, além de termos que fazem parte de uma estrutura de representação do conhecimento (taxonomias, ontologias ou tesauros), associada ao repositório. Em um segundo momento da Folksonomia Assistida (passo 5), o ambiente reconhece as informações inseridas e busca relacionamentos e associações dentro da estrutura de representação do conhecimento e do conjunto de *tags* já definidas por outros

usuários, agregando termos e devolvendo a nova lista de sugestões ao usuário. No sexto passo, o usuário atua novamente, completando o processo da Folksonomia Assistida escolhendo de forma definitiva os termos que serão utilizados na *tag* assunto. Os passos 7 e 8 correspondem a confirmação da descrição completa do recurso, a atualização dos pesos dos termos e relacionamentos de acordo com os novos elementos inseridos e a armazenagem do conjunto completo de metadados na base de dados. A cada período de tempo, o processo deverá ser avaliado por um administrador do sistema (passo 9), que pode também retroalimentar a estrutura de representação do conhecimento das áreas de especialidades, dando uma nova visão a respeito dos limites estabelecidos no domínio (SANTARÉM SEGUNDO; VIDOTTI, 2011).

Os autores afirmam que o modelo de Folksonomia Assistida não tem a intenção de limitar a criatividade do usuário nem descaracterizar o termo folksonomia, pois o usuário tem a liberdade de decidir quais os termos irá utilizar como *tags*. O processo tem como proposta primar pela “consistência das *tags*, de forma que o usuário do sistema evite abreviações, plurais/singulares ou ainda palavras que possam dificultar a recuperação da informação, posteriormente” (SANTARÉM SEGUNDO *apud* SANTARÉM SEGUNDO; VIDOTTI, 2011, p. 291).

11. Considerações finais

Com o advento da explosão informacional nas últimas décadas, principalmente quando tratamos de recursos digitais, cresce a necessidade de aperfeiçoamento das técnicas de tratamento da informação, para que esta seja descrita e recuperada com eficiência em meio ao grande repositório de recursos informacionais, a web. As bibliotecas, antes caracterizadas pelo caráter físico de suas coleções, evoluíram para o meio eletrônico, exigindo um esforço no sentido de criar novos meios de tratamento da informação contida nos recursos digitais e construir métodos alternativos para a indexação e recuperação de documentos.

A utilização de metadados não é novidade na área de Ciência da Informação, porém novas técnicas e recursos estão sendo desenvolvidos para que haja uma adaptação a essa nova realidade, onde a indexação manual de web sites, por exemplo, já é humanamente impossível. A Iniciativa Dublin Core é um exemplo de esquema de metadados que podem ser utilizados por qualquer pessoa, empresas ou especialistas, e é subsídio para a criação das ontologias, que permitem inserir semântica e relações complexas entre termos de vocabulários controlados. Além dos metadados, as linguagens de descrição de recursos como o XML e o RDF são utilizadas no desenvolvimento da Web Semântica, juntamente com padrões para codificação de ontologias como o OWL.

A partir de uma descrição dos recursos informacionais por meio de associações de termos a ontologias e o entendimento do conteúdo descrito através de inferências realizadas por agentes inteligentes, a proposta da Web Semântica é a de delegar às máquinas a tarefa de não apenas encontrar a resposta à questão levantada, como também apresentar alternativas e sugestões, baseando-se nos gostos e nos hábitos de navegação do usuário.

Existe uma variedade enorme de redes sociais disponíveis e acessadas por todos os tipos de usuários, tanto pessoas comuns quanto especialistas em determinadas áreas. O acesso a essas redes possibilita o compartilhamento de informações e a troca de experiências entre pessoas. Ao fazer parte de uma rede social, o usuário é incentivado a contribuir com o tratamento dessas informações,

por meio da folksonomia, por exemplo. É inegável que a atribuição de *tags* aos recursos digitais possui um caráter pessoal, porém pode contribuir para a identificação de hábitos coletivos e de senso comum, sendo uma alternativa para o enriquecimento de vocabulários controlados já existentes e a criação de sistemas de recuperação da informação baseados na etiquetagem coletiva.

O cenário de evolução da web também promove mudanças relacionadas à busca de informações. Os motores de busca, tradicionalmente incapazes de extrair semântica dos documentos, passam a contar com a inteligência coletiva para realizar pesquisas mais ricas, buscando e processando informações de maneira cada vez mais precisa.

Nesse contexto, a Biblioteconomia tem muito a ganhar. Sabe-se que o mínimo exigido dos agentes inteligentes da Web Semântica é saber onde encontrar a informação e ter a capacidade de recuperar, interpretar, analisar e traduzir a solicitação do usuário, disponibilizando os resultados em um formato comum. As bibliotecas digitais, dotadas das tecnologias desenvolvidas pela Web Semântica, poderão otimizar seus serviços e automatizar cada vez mais o processamento técnico de seus acervos. As novas tecnologias trazem incrementos nos índices de revocação e precisão nas pesquisas geradas pela necessidade de informação, com a redução da incidência de problemas relacionados com a polissemia e a sinonímia, considerando as informações em seus contextos de significado.

Diante de tantas transformações, o profissional da informação se vê incentivado a buscar o aperfeiçoamento em novas técnicas e saberes, no sentido de mediar e fornecer subsídios para a obtenção do conhecimento. A introdução de matérias relacionadas à ontologias e ao desenvolvimento da Web Semântica se faz necessária, pois começam a fazer parte dos pilares da Ciência da Informação. As técnicas tradicionais não deixaram de ter a sua utilidade; elas serão complementadas por novos métodos de organização da informação.

12 Referência Bibliográfica

ALVARENGA, Lídia. A teoria do conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **DataGramaZero – Revista de Ciência da Informação**, Rio de Janeiro, v. 2, n. 6, dez. 2001. Disponível em: <http://dici.ibict.br/archive/00000309/01/A_teor%C3%ADa_do_conceito_revisitada.pdf>. Acesso em: 27 dez. 2010.

AQUINO, Maria Clara. A folksonomia como hipertexto potencializador de memória coletiva: um estudo dos links e das tags no de.licio.us e no Flickr. **Liinc em Revista**, v. 4, n. 2, p. 303-320, set. 2008. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/263/175>>. Acesso em: 30 mai. 2011.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. A contribuição dos tesauros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. **Ibersid**, Espanha, p. 199-209, 2008. Disponível em: <<http://ibersid.eu/ojs/index.php/ibersid/article/view/2235/1996>>. Acesso em: 03 jan. 2011.

BRÄSCHER, Marisa. A ambigüidade na recuperação da informação. **DataGramaZero – Revista de Ciência da Informação**, Rio de Janeiro, v. 3, n. 1, fev. 2002. Disponível em: <http://www.dgz.org.br/fev02/Art_05.htm>. Acesso em: 30 mai. 2011.

CAFÉ, Lígia; MENDES, Fernanda. Estudo sobre estrutura definitória para desenvolvimento de ontologias. **Informação & Sociedade: Estudos**, João Pessoa, v. 19, n. 2, p. 71-80, mai./ago. 2009. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/3152/3026>>. Acesso em: 12 dez. 2010.

CARLAN, E.. **Ontologia e Web Semântica**. Brasília: UNB, 2006. Disponível em: <<http://eprints.rclis.org/archive/00007140/01/ECarlan.pdf>>. Acesso em: 27 Dez. 2010.

CASTRO, Fabiano Ferreira de; SANTOS, Plácida L. V. A. da Costa. Uso das tecnologias na representação descritiva: o padrão de descrição bibliográfica semântica MarcOnt Initiative nos ambientes informacionais digitais. **Ciência da Informação**, Brasília, v. 38, n. 1, p. 74-85, jan./abr. 2009. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1082/1312>>. Acesso em: 12 dez. 2010.

_____. Bibliotecas digitais: aspectos no âmbito da representação e padronização de recursos informacionais. In: CASTRO, F. F. de. **Padrões de representação e descrição de recursos informacionais em bibliotecas digitais na perspectiva da ciência da informação**: uma abordagem do MarcOnt initiative na era da web semântica. Marília: UNESP, 2008. 201 p. Dissertação (Mestrado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2008. Disponível em: <<http://dci2.ccsa.ufpb.br:8080/jspui/handle/123456789/479>>. Acesso em 21 jun. 2011.

_____. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informacionais no âmbito das bibliotecas digitais na era da web semântica. **Informação & Sociedade: Estudos**, João Pessoa, v. 17, n. 2, p. 13-19, mai./ago. 2007. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/840/1442>>. Acesso em: 12 dez. 2010.

CATARINO, Maria Elisabete; BAPTISTA, Ana Alice. Folksonomia: um novo conceito para a organização dos recursos digitais na web. **DataGramaZero – Revista de Ciência da Informação**, Rio de Janeiro, v. 8, n. 3, jun. 2007. Disponível em: <http://www.dgz.org.br/jun07/Art_04.htm>. Acesso em: 17 mai. 2011.

DZIEKANIAK, Gisele. A organização da informação e a comunicação científica: implicações para os profissionais e usuários da informação. **Em Questão**, Porto Alegre, v. 16, n. 1, p. 45-59, jan./jun. 2010. Disponível em: <<http://www.seer.ufrgs.br/index.php/EmQuestao/article/view/11434/8761>>. Acesso em: 12 dez. 2010.

DZIEKANIAK, Gisele Vasconcelos; KIRINUS, Josiane Boeira. Web semântica. **Encontros Bibli – Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 9, n. 18, 2004. Disponível em: <<http://www.journal.ufsc.br/index.php/eb/article/view/155/5471>>. Acesso em: 12 dez. 2010.

GALDO, A.; VIERA, A. F. G.; RODRIGUES, R. S. Classificação social da informação na web: tecnologia, informação e gente. **DataGramaZero – Revista de Ciência da Informação**, Rio de Janeiro, v. 10, n. 8, dez. 2009. Disponível em: <http://dgz.org.br/dez09/Art_03.htm>. Acesso em: 13 jun. 2011.

JARDIM, André Dsessards; PALAZZO, Luiz Antonio Moro. Aplicações da web semântica nas redes sociais. In: CONGRESSO NACIONAL DE AMBIENTES HIPERMÍDIA PARA APRENDIZAGEM, 4., 2009, Florianópolis. **Anais...** Florianópolis: UFSC; UCP, 2009. Disponível em: <<http://www.conahpa.org/wp-content/themes/Conahpa/papers/final108.pdf>>. Acesso em 08 jun. 2011.

JORENTE, M. J. V.; SANTOS, P. L. A. C.; VIDOTTI, S. A. B. G. Quando as webs se encontram: social e semântica – promessa de uma visão realizada? **Informação & Informação**, Londrina, v. 14, n. esp, p. 1-24, 2009. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/2215?>>. Acesso em 20 jun. 2011.

MARCONDES, Carlos Henrique. Um modelo semântico de publicações eletrônicas. **Liinc em Revista**, Rio de Janeiro, v. 7, n. 1, p. 82-103, mar. 2011. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/404/262>>. Acesso em: 21 jun. 2011.

MOURA, Maria Aparecida. Folksonomias, redes sociais e a formação para o tagging literacy: desafios para a organização da informação em ambientes colaborativos virtuais. **Informação & Informação**, v. 14, n. esp, p. 25-45, 2009. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/2196/3217>>. Acesso em: 30 mai. 2011.

_____. Informação, ferramentas ontológicas e redes sociais ad hoc: a interoperabilidade na construção de tesouros e ontologias. **Informação & Sociedade: Estudos**, João Pessoa, v. 19, n. 1, p. 59-73, jan./abr. 2009. Disponível em: <<http://dci2.ccsa.ufpb.br:8080/jspui/bitstream/123456789/628/1/2396-5184-1-PB.pdf>>. Acesso em: 30 mai. 2011.

MUCHERONI, M. L.; PAIVA, D. C. de; NETTO, M. L. Três ontologias clássicas e a web semântica. **PontodeAcesso**, Salvador, v. 3, n. 2, p. 281-298, dez. 2009. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/3621/2749>>. Acesso em: 12 dez. 2010.

NUNES, Anselmo Maciel; FILETO, Renato. Uma arquitetura para recuperação de informação baseada em semântica e sua aplicação no apoio a jurisprudência. In: ESCOLA REGIONAL DE BANCO DE DADOS – ERBD, 3., 2007, Caxias do Sul. **Anais eletrônicos...** Caxias do Sul: UCS, 2007. Disponível em: <<http://www.buscalegis.ufsc.br/arquivos/26117.pdf>>. Acesso em: 21 jun. 2011.

PEREIRA, Débora de Carvalho; CRUZ, Ruleandson do Carmo. Folksonomia e tags afetivas: comunicação e comportamento informacional no Twitter. **DataGramaZero – Revista de Ciência da Informação**, Rio de Janeiro, v. 11, n. 6, dez. 2010. Disponível em: <http://dgz.org.br/dez10/F_I_art.htm>. Acesso em: 07 fev. 2011.

PEREIRA, N. M. A.; BRITO, J. L.; SILVA, P. M. da. Perspectivas da web semântica para a biblioteconomia. In: ENCONTRO NACIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, GESTÃO E CIÊNCIA DA INFORMAÇÃO, 33., 2010, João Pessoa. Comunicação oral apresentado ao GT-01 – Biblioteconomia, tecnologias e redes sociais. João Pessoa: UFPB, 2010. Disponível em: <<http://www.slideshare.net/nairamichelle/pers-5589327>>. Acesso em: 25 jan. 2011.

PICKLER, Maria Elisa Valentim. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 65-83, jan./abr. 2007. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/viewFile/251/468>>. Acesso em: 12 dez. 2010.

RAMALHO, Rogério Aparecido Sá; FUJITA, Mariângela Spotti Lopes. A utilização de ontologias em bibliotecas digitais: um estudo metodológico. In: SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL, 2008, Niterói. **Anais...** Niterói: UFF, 2008. Disponível em: <<http://www.uff.br/ontologia/artigos/11.pdf>>. Acesso em: 03 jan. 2011.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. Web semântica: uma investigação sob o olhar da ciência da informação. **DataGramaZero – Revista de Ciência da Informação**, v. 8, n. 6, dez. 2007. Disponível em: <http://www.dgz.org.br/dez07/Art_04.htm>. Acesso em: 03 jan. 2011.

ROCHA, Rafael Port da. Metadados, web semântica, categorização automática: combinando esforços humanos e computacionais para a descoberta e uso dos recursos da web. **Em Questão**, Porto Alegre, v. 10, n. 1, p. 109-121, jan./jun. 2004. Disponível em: <<http://seer.ufrgs.br/index.php/EmQuestao/article/viewFile/86/1127>>. Acesso em: 12 dez. 2010.

SANTARÉM SEGUNDO, José Eduardo; VIDOTTI, Silvana A. B. G. Representação iterativa e folksonomia assistida para repositórios digitais. **Liinc em Revista**, Rio de Janeiro, v. 7, n. 1, p. 283-300, mar. 2011. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/414/294>>. Acesso em: 16 mai. 2011.

SAYÃO, Luis Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **TransInformação**, Campinas, v. 20, n. 2, p. 133-148, mai./ago. 2008. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=215>>. Acesso em: 12 dez. 2010.

SOUZA, Renato Rocha. Sistema de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p. 161-173, mai./ago. 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a02.pdf>>. Acesso em: 12 dez. 2010.

SOUZA, Renato Rocha; ALVARENGA, Lúcia. A web semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/50/50>>. Acesso em: 12 dez. 2010.

SPITERI, Louise F. The use of collaborative tagging in public library catalogs. In: Public Library Catalogs final report: 2006 OCLC/ALISE and Information Science Research Grant Project, February 26, 2007. Disponível em: <http://library.oclc.org/cdm4/item_viewer.php?CISOROOT=%2Fp267701coll27&CISOPTR=271>. Acesso em: 14 abr. 2011.

TRANT, J. Studying social tagging and folksonomy: a review and framework. **Journal of Digital Information**, v. 10, n. 1, 2009. Disponível em: <<http://journals.tdl.org/jodi/article/view/269>>. Acesso em: 23 mai. 2011.

VIERA, Angel Freddy Godoy; GARRIDO, Isadora dos Santos. Folksonomia como estratégia para recuperação colaborativa da informação. **DataGramaZero – Revista de Ciência da Informação**, Rio de Janeiro, v. 12, n. 2, abr. 2011. Disponível em: <http://www.dgz.org.br/abr11/Art_02.htm>. Acesso em: 09 jun. 2011.